Computational Identification of Anti-Cancer Natural Bioactive Compounds

Project report submitted to Sister Nivedita University in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE
IN
MICROBIOLOGY

Submitted By
TANMAY MAHAPATRA
Registration ID: 230010338248
Enrollment ID: 2321211001033

Advisor
Dr. Nilanjan Adhikari.
Designation: Assistant Professor.
Affiliation: Department of Pharmaceutical Technology
Jadavpur University.

School of Health Science and Translational Research
Sister Nivedita University
Kolkata

June 2025

Dr. Nilanjan Adhikari, M. Pharm., Ph. D.
ASSISTANT PROFESSOR
JADAVPUR UNIVERSITY
KOLKATA 700 032 INDIA

NATURAL SCIENCE LABORATORY
DEPARTMENT OF PHARMACEUTICAL TECHNOLOGY
E-mail: nadhikari.pharmacy@jadavpuruniversity.in
nilanjan_juphar@rediffmail.com
Mobile: +91-9433178885

Date: 05.06.2025

To whom it may concern

This is to certify that **Mr. Tanmay Mahapatra**, son of Mr. Tapas Mahapatra is studying in 2nd year in Master of Science at Sister Nivedita University, Kolkata. He has successfully completed his ***5-month internship program*** (from 1st January 2025 to 31st May 2025) under my supervision. His project work was entitled "*Computational Identification of Anti-Cancer Natural Bioactive Compounds*".

He is an outstanding, diligent, hardworking student and an excellent bench worker who can work quite well in a team as well as independently. As a person, he is honest with a pleasant personality and has the ability to learn new techniques which are required for modern research.

I wish him the best of luck in his future endeavor.

Nilanjan Adhikari

DR. NILANJAN ADHIKARI

DECLARATION BY THE STUDENT

I hereby declare that the work which is being presented in this dissertation entitled "**Computational Identification of Anti-Cancer Natural Bioactive Compounds**" in partial fulfilment of the requirements for the award of the degree of Masters of Science in Microbiology is an authentic record of my own work carried out during the period from January to May under the supervision of Dr. Nilanjan Adhikari, Department of Pharmaceutical Technology, Jadavpur University.

I acknowledge that I have read and understood the UGC regulations, 2018 (*Promotion of Academic integrity and Prevention of Plagiarism in Higher Educational Institutions*).

I confirm that this dissertation has been checked for plagiarism using the plagiarism checking software Small SEO tool. At the end of the dissertation, a copy of the summary report demonstrating similarities in content and its potential source (if any), generated online using plagiarism checking software is enclosed. I herewith confirm that the dissertation has been less than 10% similarity according to the plagiarism checking software's report and meets the UGC Regulations as well as plagiarism rules of Sister Nivedita University.

I further declare that no portion of the dissertation or its data will be published without the Institute's or supervisor's permission. I have not previously applied for any other degree or award using the topics and findings described in my dissertations.

Tanmay Mahapatra 05·06·25

(Signature of the students with Date)

Name of the Student: Tanmay Mahapatra

Registration no. 230010338248

Department: Microbiology.

SISTER NIVEDITA UNIVERSITY
সিস্টার নিবেদিতা বিশ্ববিদ্যালয়

DG Block - 1/2, Action Area I, New Town, Kolkata - 700156, West Bengal, India.

DEPARTMENT OF BIOTECHNOLOGY & DEPARTMENT OF MICROBIOLOGY
জৈব প্রযুক্তি বিভাগ ও অণুজীব বিদ্যা বিভাগ

TO WHOM IT MAY CONCERN

This is to certify that the project report titled "Computational Identification of Anti-Cancer Natural Bioactive Compounds", submitted by Mr. Tanmay Mahapatra (Registration No. 230010338248, Enrolment No. 2321211001033) to the Department of Microbiology, School of Health Science and Translational Research at Sister Nivedita University, Kolkata, is a bona fide record of research conducted by him for the partial fulfillment of the requirements for the Master of Science degree.

Date: 05.06.25

Place: Newtown, Kolkata.

Signature of Head of the Department

ACKNOWLEDGEMENT

It is my pleasure to express with a deep sense of gratitude to my master's thesis guide Dr. Nilanjan Adhikari, Assistant professor at Jadavpur University for his constant guidance, continual encouragement and understanding, more than all he taught me patience in my endeavor. My association with him is not confined to academics only but it is a great opportunity on my part to work with an intellectual expert in this field of computational biology.

I would like to express my heartfelt gratitude to Mr. Satyam Roy Chowdhury, Chancellor; Prof. (Dr.) Dhrubajyoti Chattopadhyay, Vice Chancellor; Dr. Rajat Pal, Assistant Director - Microbiology and Biotechnology, Dr. Fatema Calcuttawala, Assistant Professor and Head Microbiology for providing me the opportunity to work and for the inspirational guidance and helping me throughout the journey.

I would like to thank my parents for their support. It is indeed a pleasure to thank my friends who persuaded and encouraged me to take up and complete this task. Last but not the least, I express my gratitude and appreciation to all those who have helped me directly or indirectly towards the successful completion of the project.

Place: Kolkata. Tanmay Mahapatra

Date: 05.06.25 Signature of Candidate

LIST OF FIGURES

Fig 1: ProTox Analysis

Fig 2: Grid Box generation in PyRx window.

Fig 3: 3D diagram of Ligand-Protein (HER2) interaction.

LIST OF TABLES

Table 1: List of top ligands along with their Binding affinities and coordinates of SBD site sphere to validate docking result.

Table 2: Ligand- Protein Interaction showing interacting residues to the protein receptor.

Table 3: Toxicity analysis of lead compounds from docking in ProTox 3.0

ABBREVIATIONS

NP -Natural Product

CADD- Computer Aided Drug Design

VS- Visual Screening

ADMET- Absorption, Distribution, Metabolism, Excretion, Toxicity

AI- Artificial Intelligence

FDA- Food and Drug Administration

ML- Machine Learning

RMSD- Root Mean Square Deviation

SAR- Structure Activity Relatioship

PDB- Protein Data Bank

NCI- National Cancer Institute

PK- Pharmacokinetics

PD- Pharmacodynamics

ABSTRACT

The growing incidence of cancer worldwide calls for creative approaches to hasten the development of new treatments. This project uses in silico tools to expedite drug discovery by investigating computational methods for identifying natural bioactive compounds that have anti-cancer properties. The study evaluates natural compounds from a variety of sources, including NPASS, NPACT, and other databases for natural compounds, by combining molecular docking, virtual screening, and ADMET analysis.

Cheminformatics analyses revealed thousands of natural molecules, while virtual screening identified lead candidates with high binding affinities to HER2, oncogenic target used in this case. In this study compound such as Tanshione IIA exhibited strong binding affinities and interactions with residues at the active site of the protein, analyzed by molecular docking study.

In spite of the fact that experimental validation is necessary to verify bioactivity and safety, the study highlights the effectiveness of computational approaches in cutting down on the time and expenses involved in early-stage drug discovery. One potential avenue for improving predictive accuracy in the future is the combination of machine learning and conventional in silico methods.

This work shows how computational identification provides a strong framework to rank natural compounds for preclinical development, bridging the gap between traditional knowledge and contemporary oncology. These tactics have the potential to greatly increase the pipeline of anti-cancer drugs and meet unmet clinical needs by releasing the chemical diversity found in nature.

CONTENT

Sr. No.	Topic	Page No.
1.	Introduction	10
2.	Review of Literature	12
3.	Materials and Methods	17
4.	Results and Discussions	24
5.	Conclusion	31
6.	Appendix	32
7.	References	33

1. INTRODUCTION

The creation of aberrant cells that show fast growth, unchecked division, and the capacity to spread and invade neighbouring tissues is a defining characteristic of cancer, a general term for a broad category of diseases that can affect any part of the body.The ability to spread and invade nearby tissues is termed as metastasis, which marks the primary cause of death from cancer. Cancer is caused by mutation and epimutation of the genetic material of normal cells, which disturbs the normal balance between cell proliferation and death, defying the normal boundaries of cell division (Pantel, Klaus, et. al. 2009). Mutations responsible for causing malignancy can be acquired gradually throughout the lifetime or can also have hereditary origins such as retinoblastoma. Mutations in the proto-oncogenes involved in cellular growth processes and tumor suppressor genes, frequently result in cancer. For instance, Retinoblastoma, which is a form of retinal cancer caused due to mutations in the RB1 gene, which is a tumor suppressor gene. Apart from gene mutations various physical and chemical substances causing cancer, termed as carcinogens also contribute to the development of cancer. Physical causes include ultraviolet rays, ionizing radiation such as X-ray, gamma rays etc. while chemical carcinogens include tobacco smoke, asbestos and certain industrial pollutants.

Cancer is a leading cause of death worldwide, accounting for nearly 10 million deaths in 2020, according to official databases of the World Health Organization (WHO). The most common cancers, in terms of new cases as observed in 2020 are breast cancer accounting for 2.26 million cases, lung cancer (2.21 million cases), colo-rectal cancer (1.93 million cases), prostate cancer (1.41 million cases) (Torre, Lindsey A., et al. 2016). The incidence of cancer is observed to rise dramatically with age due to increase in risk factors and decrease in efficiency of cellular repair mechanisms.

Till date there are various modes of cancer treatments, ranging from surgical methods to chemotherapy, radiation therapy, immunotherapy and so on. Although there is a significant advancement in cancer therapies there are also potential side effects, toxicity causing harm to the patient, and limitations in targeting specific cancers. The side effects are due to cancer therapies that target cancer cells all while destroying normal healthy cells in the process. Common short term side effects include nausea, vomiting, anemia, hair loss, diarrhea, increased susceptibility to infections and so on, whereas long-term effects comprise bone

(osteoporosis) and joint problems, cognitive issues, defects in digestive and nervous systems (neuropathy), reproductive systems (early menopause and infertility).

To counter this wide range of adverse side effects alternative treatments are needed to improve the survival and quality of life of patients, while maintaining the efficacy of treatment. Natural products isolated from various plants and marine organisms provides an alternative solution to cancer treatment. Natural products while providing nutritional value are also being explored for their therapeutic potential (Huang, M. et. al. 2021). Some classes of compounds with anti-cancer properties are Terpenoids, Flavonoids, Alkaloids, Lignans.

Natural compounds display a variety of anti-tumor activities from promoting apoptosis, inhibiting angiogenesis, anti-proliferation activity, inhibiting various cell signaling pathways, to regulating immune function. Natural compounds triggers cell arrest at specific cell cycle phases (G2/M, G0/G1) by targeting cyclins and cyclin dependent kinases (CDKs). Various cancer associated cell signaling pathways such as Wnt/b-catenin, JAK/STAT (Janus kinase/ Signal transducer and activator of transcription) pathway, phosphoinositide 3-kinases (PI3K)/Akt/mammalian target of rapamycin (mTOR) pathway are also targeted suppressing cancer progression (Nan, Yang et al. 2023)

2. REVIEW OF LITERATURE

2.1. Bioinformatics and structural elucidation

Identification of natural compounds is made extremely difficult by their structural complexity. The process has accelerated thanks to the use of computer-assisted structural elucidation (CASE) tools in conjunction with spectroscopic methods like mass spectrometry and NMR. The NP-Structure predictor algorithm, for example, compares scaffolds and side-chain combinations using LC-MS data to produce possible molecular structures (Harn et al., 2017). Hyphenated methods like GC-MS/MS and LC-MS/MS improve resolution by decreasing overlapping m/z ratios and increasing the precision of constituent identification in natural mixtures (Rinschen et al., 2019). By combining topological constraints and spectral data, fuzzy logic-based systems make it possible to clarify large molecules (>100 atoms). By ensuring that novel compounds are given priority for additional research, these tools are essential for dereplication.

2.2. Molecular modeling and virtual screening

In the search for anticancer drugs, computational screening has emerged as a key component. To forecast how natural compounds will interact with cancer targets, molecular docking and dynamics simulations are frequently employed. According to, (Iqbal et al. 2022), molecular docking studies revealed that α-pinene and D-limonene from Foeniculum vulgare had strong binding affinities (−6.0 and −5.9 kcal/mol, respectively) to the estrogen receptor α (PDB ID: 6CHZ), a target for breast cancer. Similar to this, terpenoids from Cladosporium species were screened against the estrogen receptor α using density functional theory (DFT) and integrated molecular dynamics (MD) calculations, which revealed electronic properties and stable binding modes that are essential for inhibition (Alsuhaibani et al., 2022).

In order to create inhibitors, fragment-based drug discovery (FBDD) makes use of tiny molecular fragments. For example, FBDD-identified pyrollopyridine fragments resulted in the development of vemurafenib which is an FDA approved anticancer drug (Liu & Quinn, 2019). Further predictive accuracy is enhanced by machine learning models employing QSAR (Quantitative structure- activity relationships) and ADMET (Absorption, distribution, metabolism, excretion, toxicity) properties.

2.3. Computational Vs Traditional methods

Computational science is a field of applied biology that applies the ways from computer intelligence, arithmetic, measurements, and information to measure normal frameworks, demonstration & associations. computational styles comparable as calculations, fine models, and reenactments to test huge and complex characteristic information sets, decipher test comes about, and actuate unused common sapience. Conventional biology contained item like web analysis, engineered chemistry, ethnopharmacology (process of using plant & animal-based product for medicinal purpose), trial and error, pharmacology etc.

Computational science and traditional study in cancer investigation vary unnaturally in their approaches, ways, and operation. Conventional Mold gives coordinate unthinking substantiation (e.g., protein- DNA relations). Set up gold standards for FDA medication favouring. But fizzled for Destitute versatility for examining excrescence differing qualities. Computational design captures cellular differences and dynamicity more efficiently compared to traditional approach. By demonstrating person cells, computational models can specifically join exploratory discoveries into reenactment rules. This empowers the consider of how particular cellular characteristics (e.g., relocation rate, division concordance, multiplication inevitability) impact excrescence development and cure reaction Computational think about connect microenvironmental variables. With progresses in computing, these models can imagine millions of cells, making it conceivable to fleetly investigate a wide run of common scripts and pre-screen medicinal techniques some time recently moving to costly and time-expending research facility trials. Integration with exploratory framework models can handle frameworks of a scale comparable to in vitro and without a doubt in vivo trial, permitting for coordinate comparison and affirmation.

Computational science offers eminent taken a toll edge compared to conventional exploratory styles, particularly in ranges like pharmaceutical revelation, cancer investigation, and expansive- scale common information investigation. In silico analyses huge- scale, tall- outturn information (genomics, proteomics), Can ponder whole genomes, systems, or populaces contemporaneously but conventional design handles lower datasets, as often as possible created physically.

2.4. HER2 Family

HER- 2 (also known as HER2 or ErbB2) is a 185- kDa protein receptor with tyrosine kinase effort and broad homology to the epidermal development calculate(EGF) receptor. HER- 2/ neu is dependable for generally 20 – 25 of all ovarian and bone cancers, 35 – 45 of all pancreatic adenocarcinomas, and over to 90 of colorectal lymphomas. HER- 2/ neu overexpression speaks to a marker of destitute prognostic. HER- 2/ neu-positive excrescence cells are possibly great targets for excrescence- receptive cytotoxic T lymphocytes which have been utilized in immunotherapeutic trials. In expansion, the "humanized" monoclonal counter acting agent Herceptin has been tried in a few clinical trials and demonstrated to be a compelling adjuvant cure for HER- 2/ neu-positive bone and ovarian cancers.

Overexpression of the mortal epidermal development calculate- 2(HER2) oncogene in mortal bone lymphomas has been related with a more forceful course of complaint. The reason for this affiliation is still vague, in spite of the fact that it has been proposed to rest in expanded expansion, vessel adaptation, and/ or invasiveness. On the other hand, prognostic may not be straightforwardly related to the nearness of the oncoprotein on the cell film, but or maybe to the bone melanoma subset connected by HER2 overexpression and characterized by a impossible to miss quality expression profile. The effective improvement of trastuzumab (Herceptin), ananti-HER2 counter acting agent, has had a major effect on the treatment of bone cancer. Since moreover, an assortment of HER2 particular antibodies and little atomic hindrances have been surveyed in clinical trials.

Both HER2 receptor and the entirety HER2 signaling pathway have been impressively considered as the target for cancer cure. Both the upstream and downstream of the HER2 signaling pathway are promising targets to square the flag and hinder the excrescence development. too, as an overexpressed cyto- film protein, HER2 is an tempting marker for focused on pharmaceutical conveyance to excrescence cells. Due to the critical parallels in the protein arrangement between the EGFR(HER1) and the item of the neu quality, it was named HER2 and proposed to play a portion as a kinase receptor at the cell confront. In 1986, Drebin et al. illustrated the tumorigenic projection of NIH3T3 cells, when transfected with the HER2 quality in vivo.

2.5. Computational Success of Natual Compounds

2.5.1. Taxol

Taxol is a most famed & controversial chemo- remedial medicine(deduced from pacific yew tree), also known as paclitaxel, set up in the 1960s, which is responsible foranti-mitotic effect on oncogenic cells and well known for the treatment of colorful types of cancer like bone, ovarian, lung cancers etc. It's celebrated because of its efficacity as an anticancer medicine and controversial because the medicine and the yew tree came bogged in several sensitive political issues with broad public policy recrimination. Taxol is resolved as a rescuer of cancer cases. It's the best- dealing cancer medicine. The medium of Taxol is to stabilize and help microtubules depolymerization (assembly or disassembly of microtubules by the addition or junking of tubulin heterodimers from a microtubule), responsible for cell cycle arrest at the G2/ M phase and cell death. in mitosis cell division, Taxol also intervenes cells survival through apoptotic- regulating proteins, Bcl- 2anti-apoptotic protein family and also other signaling pathways like JNK/ SAPK, NFκB pathway. Taxol's activates or inhibits either thepro-apoptotic (BAX/ BAK) oranti-apoptotic (Bcl- 2/ MCL- 1) Bcl- 2 proteins to beget cell death in the cancer cells. Calcium stored in endoplasmic reticulum also could be attacked by Taxol and be released to disrupt the calcium homeostasis (regulation of calcium in blood) to disrupt the cell-survival.

Despite its wide use, Taxol has certain disadvantages. The main challenges facing Taxol are the need to find an environmentally sustainable product system grounded on the use of microorganisms, increase its bioavailability without plying adverse goods on the health of cases and minimize the resistance presented by a high chance of cells treated with paclitaxel.

2.5.2. Vincristine

vincristine is a natural life rescuer emulsion, contained medicinal value, insulated from Madagascar periwinkle factory, also known as Vinica rosea. Vincristine medicines contained strong inhibitory effect on monozygotic leukemia, testicular cancer, bone cancer, liver cancer, ovarian cancer, head and neck cancer, solid sarcoma, and nasty melanin in a variety of robotic or transplanted lymphocytic leukemia. Vincristine specifically interferes with cancer cell division by binding with tubulin, which is an essential protein for forming microtubule during cell-division.

This list help tubulin dimers from polymerizing into microtubule assembly, which are essential for forming mitotic spindle during cell division. Chromosome isolation is blocked during metaphase. This leads to mitotic arrest, where cell division stalled and this is directed to apoptosis. Vincristine also convinced to tone association of tubulin into abnormal helical aggregation which dismembering microtubule network.

It may intrude with other cellular process, similar as amino acid, cyclic AMP, also glutathionine metabolism, calmodulin dependent calcium signalling and also anticancer. Vincristine must n't be given intrathecally as can be fatal. Vincristine induce nervous complaint as a result of over use.

2.5.3. Curcumin

Curcumin, the primary bioactive emulsion in turmeric (Curcuma longa), has been considerably studied for its implicit part in cancer treatment. Inhibit the cancer cell proliferation, irruption, metastasis, angiogenesis and apoptosis suppress inflammation, and modulate colorful molecular targets involved in excrescence growth and metastasis.
Clinical trials have estimated curcumin in a range of cancers, including pancreatic, bone, colorectal, prostate, and multiple myeloma. For illustration, in a phase 2 clinical trial involving cases with advanced pancreatic cancer, oral curcumin (8 g daily) led to natural exertion in some cases, including stable complaint and significant excrescence retrogression. In bone cancer, combining curcumin with standard chemotherapy(docetaxel) was set up to be effective and safe in advanced cases.

Studies in colorectal cancer have shown that curcumin, when administered alongside chemotherapy, is well- permitted and may enhance treatment efficacity.
Curcumin has also been used as an adjuvant remedy, perfecting quality of life and reducing systemic inflammation in cases witnessing standard chemotherapy for solid excrescences. also, curcumin has shown implicit in easing side goods of cancer treatments, similar as oral mucositis and radiation dermatitis.
Despite its promising goods, curcumin's clinical operation is limited by its low bioavailability. To address this, colorful phrasings and delivery systems are being developed to enhance its immersion and remedial impact.

3. MATERIALS & METHODS

3.1. Materials

3.1.1. PyMOL

DeLano Scientific LLC was the first company to commercialize Warren Lyford DeLano's open-source molecular visualization system, PyMOL. Later, in 2010, PyMOL was purchased by Schrodinger Inc., which assumed full responsibility for the software's development, upkeep, and marketing. According to (Yuan et al. 2017), PyMOL is currently one of the most popular 3D macromolecular visualization tools. The majority of the common representations for macromolecular structures, including wire bonds, cylinders, spheres, ball-and-stick, dot surfaces, solid surfaces, wire mesh surfaces, backbone ribbons, and cartoon ribbons, can be produced by PyMOL in high quality images.

PyMOL is written in Python, which is one of the most popular programming languages allowing it with features and expandability conditions unmatched by any traditional packages. PyMOL is also capable of editing a macromolecule and analyzing Protein-Ligand interactions, molecular simulations (MS). (Delano et. al. 2002). PyMOL features cross-platform compatibility, and both a graphical user interface (GUI) and a powerful command line script, providing accessibility for both novice to expert researchers. PyMOL has become an essential tool in structural biology, bioinformatics, drug discovery.

3.1.2. Discovery Studio

Dassault Systems BIOVIA developed and marketed Discovery Studio, a suite of software tools for simulating systems of small molecules and macromolecules. The suite uses software algorithms like CHARMM, MODELLER, DELPHI, ZDOCK, and others that were initially created in the scientific community. To assess the binding free energy, binding affinity, and other characteristics of the ligand-target complex, Discovery Studio offers a number of analysis tools. Creating high-quality figures for publication is made simple by the software's user-friendly interface, which lets users alter the look of the images and animations (Baroroh, Umi, et al. 2023).

The sequence analysis in discovery studio is done by using BLAST (Basic Local Alignment Search Tool) and protein modeling by DS Modeller. Discovery studio can be operated in both Linux and Windows based environment. (Sisodiya, Pandey et.al. 2012). Discovery studio

offers an integrated platform with a wide range of molecular modelling and simulation tools also providing users to customize their visualization output.

3.1.3. PyRx

PyRx is a virtual screening software for computational drug discovery that can be used to dock small – molecule libraries to a macromolecule to find the lead compounds with desired biological function. It is an open-source software with a easy to use interface that supports all major operating systems (Windows, Linux, and MacOS) (Dallakayan, Sargis, et.al. 2014). PyRx performs Docking based virtual screening (DBVS) methods to explore the behavior of small molecules in the binding region of targets. With PyRx, medicinal chemists can conduct virtual screening from data preparation to job submission and analysis. PyRx is a useful tool for computer-aided drug design (CADD) because it has a docking wizard and an intuitive user interface. Additionally, PyRx has a strong visualization engine and chemical spreadsheet-like functionality, both of which are critical for logical drug design. (Ounthaisong, Untiwa, et. al. 2017).

Autodock 4 and Autodock Vina are two of the many open-source docking programs that PyRx uses. To create input files, use AutoDock Tools. Python is a language used for programming and scripting. Kitware's Visualization Toolkit (VTK). To import SDF files, eliminate salts, and minimize energy usage, launch Babel. For 2D plotting, use Matplotlib.

3.1.4. AutoDock Vina

Oleg Trott and Andrej J. Kosztin created Autodock Vina in 2009 as a quicker and more precise substitute for the Autodock program (AutoDock 4). When compared to AutoDock 4, Autodock Vina achieves a speedup of about two orders of magnitude and greatly increases the binding mode prediction accuracy. In a manner that is clear to the user, AutoDock Vina also computes the grid maps and groups the results automatically. (Olson, Trott, and others, 2010). Vina employs a gradient-based local search genetic algorithm, a global optimization technique, to forecast how small molecules will bind to their protein targets. Using random ligand conformations, the algorithm refines binding energies by searching and perturbing coordinates exhaustively, ultimately enabling global minimum energy identification.

The algorithm uses random ligand conformations, which then exhaustively searches and perturbs coordinates to refine binding energies finally enabling global minimum energy

identification (Sarkar, Arkadeep, et. al. 2024). In multicore servers, Vina uses multithreading optimally to increase the speed of docking simulations, reported to achieve a 60-fold increase in speed with similar or better accuracy compared to Autodock 4 (Biesiada, Jacek, et. al. 2011).

3.1.5. Open Babel

Open Babel is an open-source chemical toolbox designed to speak the many languages of chemical data. It's an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas. Open Babel interconverts over 110 formats. To represent such a wide variety of chemical and molecular data, a library is required that implements a wide range of cheminformatics algorithms, from partial charge assignment and aromaticity detection, to bond order perception and canonicalization (O'Boyle. Noel M., et. al. 2011). Open Babel is written in C++ with Python, Perl, Java, Ruby, R, PHP, and Scala bindings. Its capabilities include interconversion of chemical compounds of more than 110 formats, fingerprint generation, property determination, similarity and substructure search, and molecular force fields. It can also produce 3D structures through the systematic enumeration of torsions and energy minimization (Pirhadi, Somayeh, et. al. 2016)

3.1.6. SwissADME

SwissADME is a web tool providing free access to a pool of predictive models for physicochemical properties, pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules (Daina, Michielin, et. al.). SwissADME adapts software applications to interpret properties like ADME (Absorption, Distribution, Metabolism and Excretion) of small compounds. It was designed in a way making it availability easy to even the most general category of scientific community and its user-friendly interface aids even non-experts in Computer-Aided Drug Discover (CADD). SwissADME is also cost-free making it easily available for novice scholars for carrying out in-silico studies quickly and efficiently (Bakchi, Bulti, et. al. 2022).

Other significant features are the BOILED-Egg model (predicts passive intestinal absorption and Blood-brain barrier (BBB) permeation, iLOGP and Bioavailability Radar (graphical representation of drug-likeliness parameters), Multiple predictive models.

3.1.7. ProTox

ProTox 3.0 is a virtual lab for the prediction of toxicities of small molecules, which is an important part of the drug design development process. ProTox 3.0 incorporates molecular similarity and machine learning models to predict 61 toxicity parameters such as acute toxicity, organ toxicity, clinical toxicity and molecular-initiating events (MOE), adverse outcome pathways (Tox21), several other toxicological endpoints and toxicity off-targets. ProTox 3.0 webserver is a freely available platform for in-silico toxicity prediction, used by toxicologists, regulatory agencies, computational chemists, and medicinal chemists. The webserver generally takes a chemical 2D structure as input to predict the toxicological profile with a confidence score of the compound that can be easily downloaded (Banerjee, Kemmler, et.al. 2024). ProTox 3.0 data is stored in a relational MySQL database. To handle the chemical information within the database, the MyChem package is used. For most of its functions, MyChem relies on the Open Babel toolbox. For advanced users, data can be queried using a simple POST interface with a suitable programming language. To run the script, generally Python (3.6 or newer) version in installed. Computational toxicological predictions with ProTox 3.0 can help to reduce the amount of animal experiments and save animal lives.

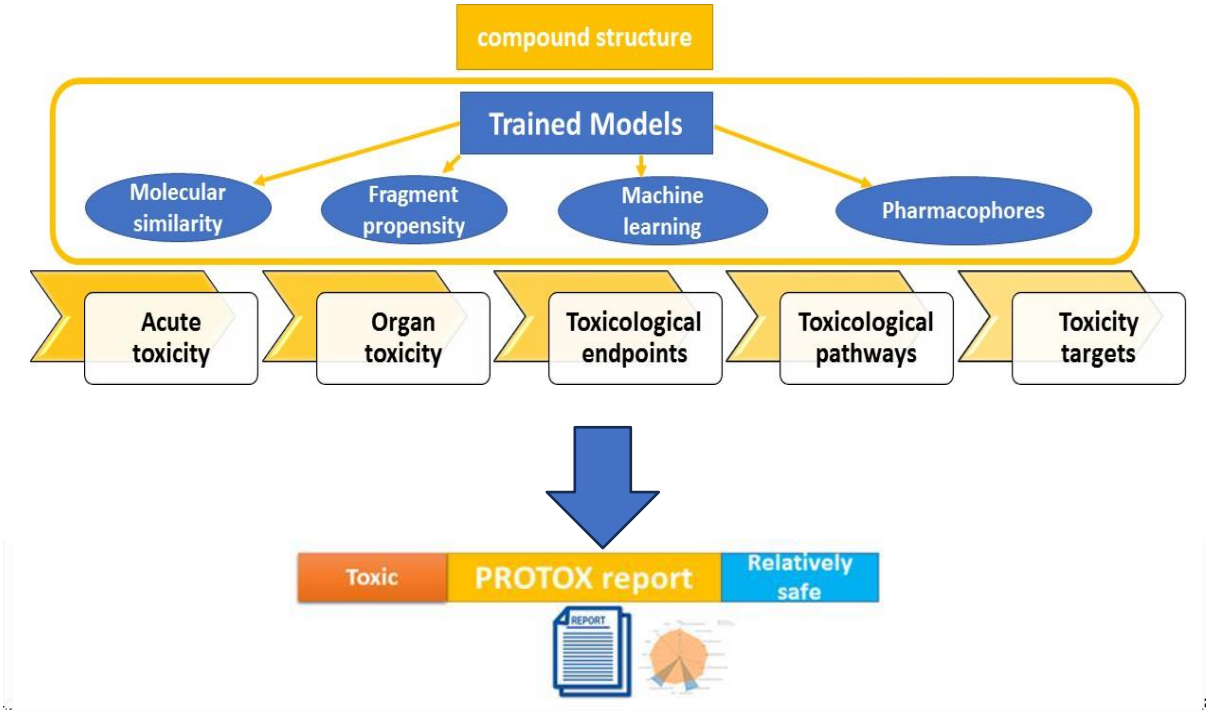


Fig 1: ProTox Analysis

3.2 Methods

3.2.1. Selection of Target Protein

After careful consideration from different literature sources, comprising review articles HER2 (Human epidermal growth factor receptor 2) was selected for docking studies. The crystal structure of kinase domain of HER2 was obtained from RCSB PDB (Protein Data Bank) having PDB ID: 3PP0 . The structure was downloaded in .pdb file format from the website (https://www.rcsb.org). The HER2 crystal structure was obtained from the website which is a high-resolution structure (2.25 Å) containing a selective inhibitor (Aertgeerts, Kathleen et. al. 2011).

3.2.2. Selection of Ligands

The chemical structures of natural compounds were obtained from NPASS (Natural product activity & species source) database (https://bidd.group/NPASS/) and NPACT (Naturally occurring plant based anticancerous compound-activity-target) database (https://webs.iiitd.edu.in/raghava/npact/browse.php). NPASS is a leading data source for the natural product research community since it was first released in 2017. NPACT is also one such cumulated database of plant derived natural compounds that exhibit anti-cancerous activity containing 1574 entries. NPACT also provides bioactivities of natural compounds against different cancer cell lines and their molecular targets.

A total of 423 natural compounds were obtained which was later screened based on various parameters. The chemical structures of ligands were obtained in SDF format.

3.2.3. ADME Screening

The properties of Absorption, Distribution, Metabolism, Excretion (ADME) of 423 natural compounds were carried out to investigate how they are processed by a living organism. This was done using the Swiss ADME web server (http://www.swissadme.ch/). Swiss ADME web server allows to predict ADME parameters, compute physiochemical descriptors, pharmacokinetic properties, and druglike nature, medicinal chemistry friendliness (Daina, Antoine et. al. 2017)

The properties investigated involved the molecular weight, lipophilicity (MLOGP, XLOGP3 and WLOGP values), water solubility (with the ESOL logS and Ali logS models), polarity (Topological Polar Surface Area – TPSA), the number of hydrogen bond acceptors and donors, molecular flexibility (rotatable bonds), synthetic accessibility, number of heavy atoms, fraction of carbon atoms in sp^3 hybridisation (fraction-Csp^3), GI absorption, and BBB permeability.

The assessment also involved major drug-likeliness rules like Lipinski (Lipinski, et. al. 2004), Ghose (Ghose et. al. 1999), Verber, Egan and Muegge filters and also Bioavailability score. The parameters were later examined carefully by applying filters to take into account acceptable values of natural compounds that can be suited as drug candidates. A total of 34 natural compounds was curated from a list of 423 natural compounds.

3.2.4. Protein Preparation

The HER2 (PDB ID: 3PP0) protein was prepared for molecular docking through a series of steps. The protein structure was imported in Discovery Studio Visualizer, in which bound water molecules are removed. The pre-bound crystallized ligand present in the structure of HER2 (PDB: 3PP0) was extracted and saved as control, revealing the active site of the protein, which is used to form the SBD site sphere around the active site. The coordinates of the active site are used to form the grid, later in the docking process. Polar hydrogens are also added to the macromolecular structure.

Then different charges were assigned to the extracted macromolecular structure (protein) such as Gasteiger charges, Kollman charges, in AutoDock Tools (Dar, Ayaz M., et al. 2017). The structure was saved in PDBQT file format which is the required format for docking.

The process of identification of the active site in a protein is critical when the co-crystallized ligand is absent in the crystal structure. For this we can perform comprehensive literature review of the source papers to identify the active site of residues. If any established drug giving the same pharmacological action of interest is available for the protein, then the active sites for that drug can be identified. Another method is that the inbuilt algorithm in visualization and docking software is used to identify the active site of the protein.

3.2.5. Molecular Docking

Molecular docking methods are routinely used for prediction of protein–ligand interactions and provides opportunity in selecting potent inhibitor molecules as a part of virtual screening of

large databases (Elokely, Khaled, et. al. 2013). For the docking purpose PyRx, which is an open-source software for Computer-Aided Drug Design (CADD) was used to perform multi-ligand docking on HER2 protein (PDB ID: 3PP0).

PyRx consists of an integrated Open Babel interface. The Open Babel interface was used to add the 34 curated natural compounds for energy minimization and later converted to PDBQT format (supported file format) for docking. Under the macromolecule section the protein structure was imported. After the minimization process is done, the grid resolution was centered at 17.098971x16.548676x26.600647 along x, y, and z-axes, respectively, for a grid size of 40x30x40 Å to defined the binding site. The docking of all the ligands to the target protein was carried out in AutoDock Vina, according to the strict protocol. AutoDock Vina is also integrated in the PyRx software negating the need for generating of a configuration file (.txt) file, also the used of Cmd line script is not necessary. The docking was run with exhaustiveness set at 8. The results generated were later analysed. The cocrystallized ligand that was already previously bound to the protein structure, serves as the default and was first docked to the binding site of HER2, and the resulting interaction was compared to that of the 34 ligands docked in similar active sites using the same grid box dimension.

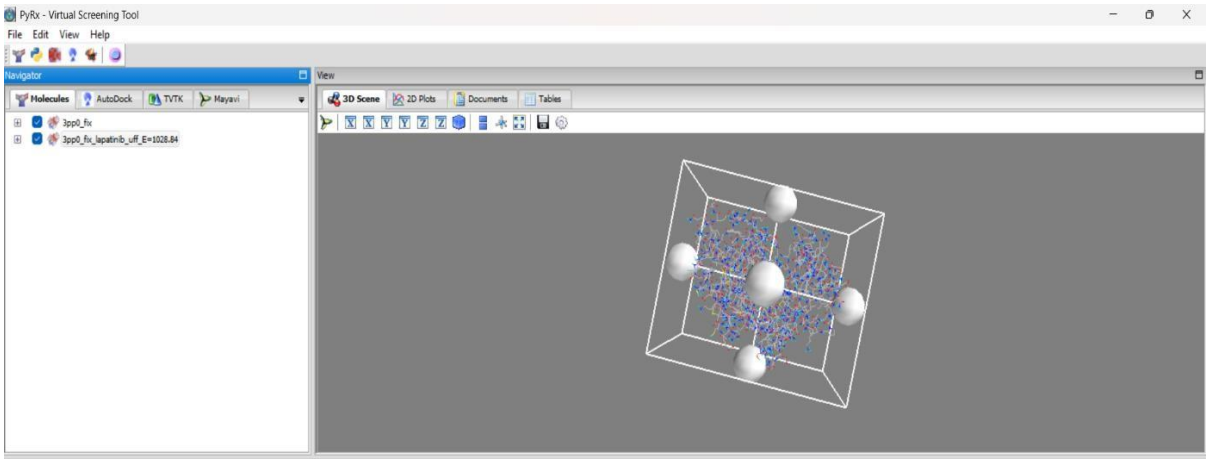


Fig 2: Grid Box generation in PyRx window.

4. Results and Discussions

Human Epidermal Growth Factor Receptor type 2 (HER2), is a protein encoded by ERBB2 gene, belonging to the family of oncogenic proteins whose inhibition of amplification or over-expression has been shown to be cancer causing.

4.1. Results

4.1.1. Filtering dataset to form a curated list of druglike NPs

In this study, filters were used on physiological properties of compounds that were generated using Swiss ADME to isolate druglike NPs (Natural Products). The properties involved molecular weight, lipophilicity (MLOGP, XLOGP3 and WLOGP values), water solubility (with the ESOL logS and Ali logS models), polarity (Topological Polar Surface Area – TPSA), the number of hydrogen bond acceptors and donors, molecular flexibility (rotatable bonds), synthetic accessibility, number of heavy atoms, fraction of carbon atoms in sp^3 hybridisation (fraction-Csp^3), GI absorption, and BBB permeability.

The assessment also involved major drug-likeliness rules like Lipinski, Ghose, Verber, Egan and Muegge filters and Bioavailability score. This generated a total of 34 compounds from a list of 422 compounds taken for the study.

4.1.2. Analysis of Docking results

The binding affinities of NPs against the target protein HER2 (PDB 3PP0) were computed with the help of AutoDock Vina. A predetermined number of models are generated for each ligand which are different poses with different binding affinities. These poses are different ways by which a ligand can fit into the active site of the receptor, in this case HER2. The output was then arranged in order of its binding energies (Kcal/mol), and RMSD value. Lower the value of RMSD higher is its docking accuracy. It indicates a better match between the docked and reference poses, giving a more accurate prediction. The output structure files were saved for further analysis of ligand- protein interactions

Table 1: List of top ligands along with their Binding affinities and Coordinates of SBD site sphere to validate the docking result.

Sr. No.	Ligand (PubChem ID)	Binding Affinity (Kcal/mol)	SBD site sphere coordinates (x, y, z) axis
1.	5154	-11.1	15.150640 13.457920 27.917920
2.	65752	-10.9	17.041750 15.424083 26.869292
3.	114917	-10.4	15.137524 11.893190 28.889762
4.	164676	-10	15.425045 13.025591 28.250545
5.	10680	-9.8	16.740118 16.932235 25.772588
6.	5281787	-9.7	17.655870 15.508261 27.202522
7.	6603789	-9.4	15.667522 16.521783 26.894174
8.	11349	-9.1	17.040421 17.044842 25.536947
9.	Control	-8.6	17.098971 16.548676 26.600647

4.1.3. Ligand-Protein Interactions

The structural validation of the docking procedure is done by extracting the co-crystallized ligand from the active site of the receptor. Then output files of PyRx docking procedure were interacted with the active site of the protein and studying the ligand- protein interaction in Discovery Studio Visualizer.

The Ligand-protein interactions are observed by accessing the Ligand- Protein Interaction panel in Discovery Studio. There are various kinds of interactions by which a ligand binds to its target: Electrostatic, which are mostly charged interactions. Hydrogen bonds, these are dipole-dipole interactions and are the most important ones. Hydrophobic interactions in which hydrophobic groups of ligands occupy the hydrophobic pockets of the target protein. There is

also a different type of very strong interactions with metallic cations. From the given data in the Table 2. it was concluded that the interacting residues with the selected ligands is similar to that of the inbound co-crystallized ligand.

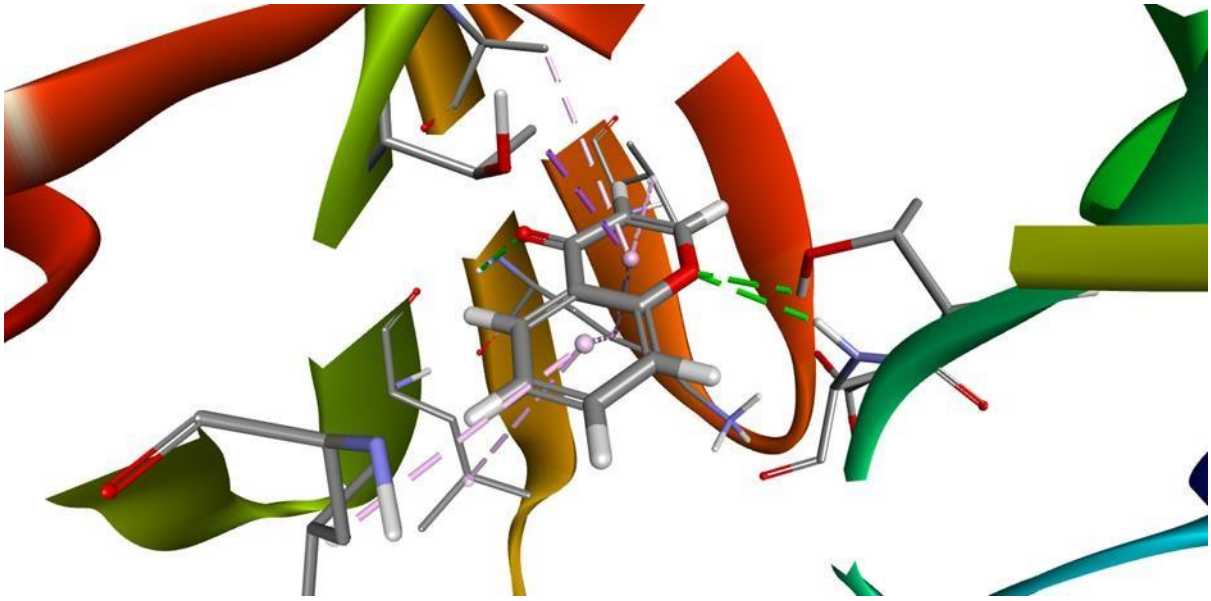
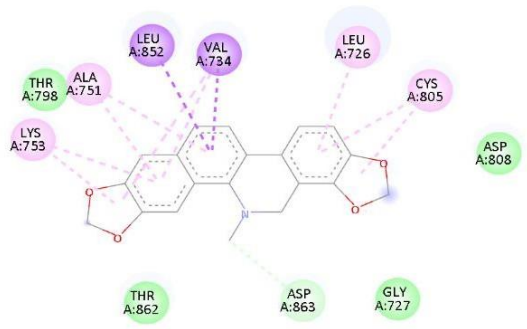
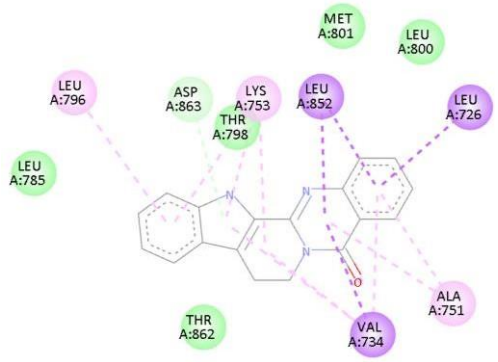
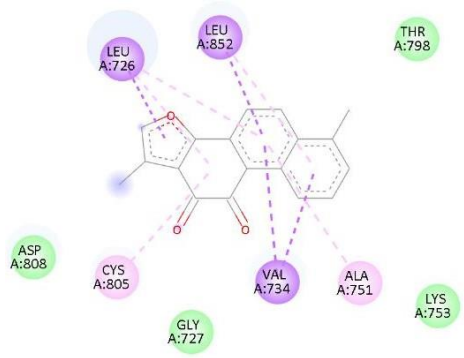
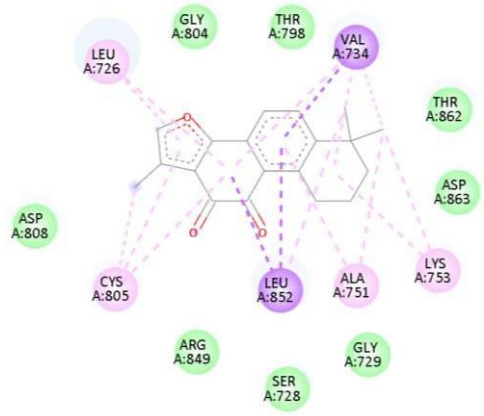
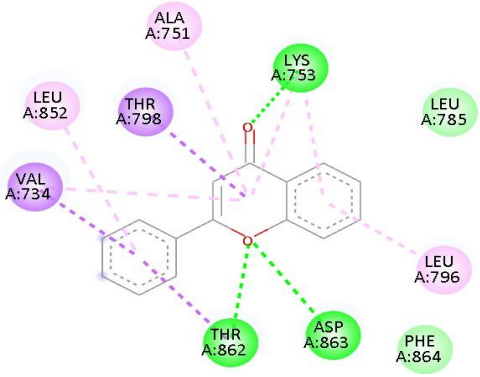
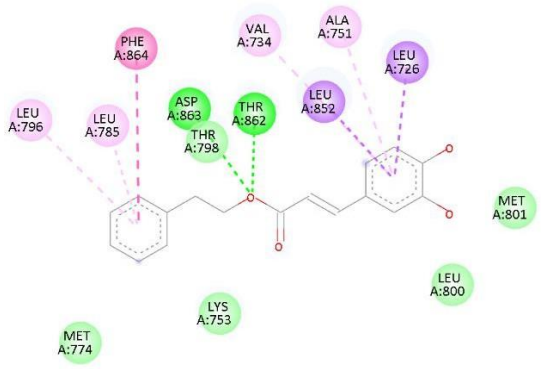
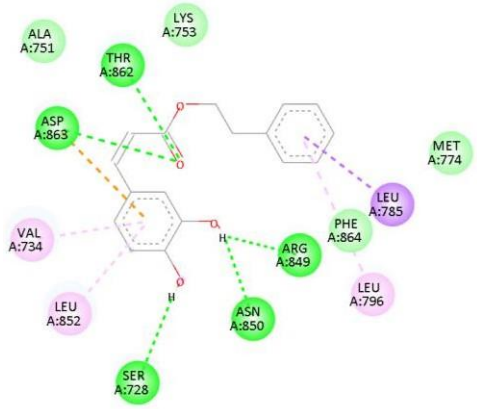
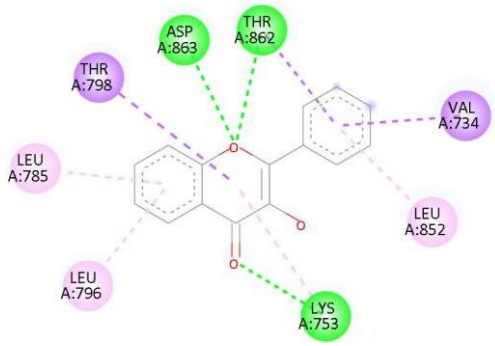
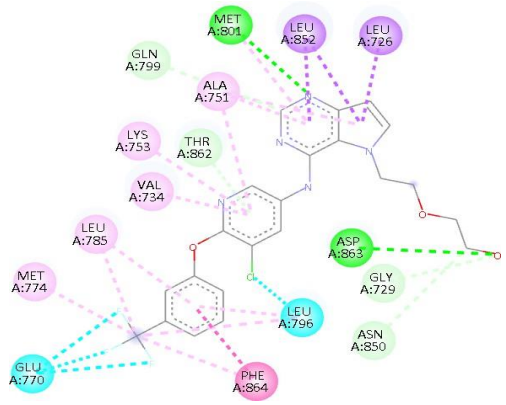


Fig 3: 3D diagram of Ligand-Protein (HER2) Interaction

Table 2: Ligand- Protein Interaction showing interacting residues to the protein receptor

Ligand Name	**2D Interaction Diagram**	**Interacting Residues**
5154		LYS 753, ASP 863 ALA 751 LEU 852 VAL 734 LEU 726 CYS 805

65752		LEU 769, THR 798, LYS 753, LEU 852, LEU 726, ALA 751, VAL 734, THR 862, LEU 785
114917		LEU 726, LEU 852, ALA 751, VAL 734, CYS 805, ASP 808, THR 798, LYS 753
164676		LEU 726, VAL 734, LYS 753, ALA 751, LEU 852, CYS 805, THR 798, ARG 849, GLY 729, ASP 863. THR 862
10680		VAL 734, LEU 852, THR 798, ALA 751, LYS 753, LEU 796, ASP 863, THR 862, LEU 785, PHE 864

5281787		LEU 796, LEU 785, PHE 864, ASP 863, THR 798, THR 862, VAL 734, LEU 852, ALA 751, LEU 726, MET 801, LEU 800, LYS 753, MET 774
6603789		LYS 753, THR 862, ASP 863, VAL 734, LEU 852, SER 728, ASN 850, ARG 849, PHE 864, LEU 785, LEU 796, MET 774, ALA 751
11349		LEU 796, LEU 785, THR 798, ASP 863, THR 862, LYS 753, LEU 852, VAL 734
Control		GLU 770, MET 774, LEU 785, VAL734, LYS 753, THR 862, ALA 751, GLN 799, MET 801, LEU 852, LEU 726, PHE 864, LEU 796, ASP 863, GLY 729, ASN 850

4.1.4. Toxicity Analysis

Compound 11349 was selected based on its favourable toxicity (Table 3) profile compared to other compounds in the dataset. When compared to the other compounds in the dataset, PubChem Id: 11349 was judged the best from a toxicity standpoint. It has a predicted LD50 value of 1000 mg/kg, placing it in toxicity class 4, which fits well within an acceptable range for many therapeutic agents. More importantly, it shows a negative result for hepatotoxicity, nephrotoxicity, respiratory toxicity, cardiotoxicity, carcinogenicity, and mutagenicity, which are the major concerns regarding major organ damages or genetic damages. In summary, PubChem ID:11349 is a well-balanced compound with respect to any potential toxicity which makes it a viable candidate for drug development.

Table 3: Toxicity analysis of lead compounds from docking in ProTox 3.0

PubChem ID	*Predicted LD50 (mg/kg)*	*Predicted Toxicity Class*	*Hepato Toxicity*	*Nephro Toxicity*	*Respiratory Toxicity*	*Cardio Toxicty*	*Carcino genecity*	*Muta genecity*
164676	1230	4	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive
114917	1655	4	Inactive	Inactive	Active	Inactive	Inactive	Inactive
11349	2500	5	Inactive	Active	Active	Inactive	Inactive	Inactive
6603789	5000	5	Inactive	Active	Inactive	Inactive	Inactive	Inactive
5281787	5000	5	Inactive	Active	Inactive	Inactive	Inactive	Inactive
10680	2500	5	Inactive	Inactive	Inactive	Active	Active	Inactive
Control	500	4	Inactive	Active	Active	Inactive	Inactive	Inactive

4.2. DISCUSSIONS

Validating docking data is an essential step that makes sure that the ligand has bound precisely and correctly inside the protein's active pocket, which verifies the accuracy of the docking study. The protein used for the study already had an inbound co-crystallized ligand inhibitor. This ligand was used for generating SBD site sphere revealing the active pocket of the protein. The inbound ligand namely **03Q** (2-{2-[4-({5-chloro-6-[3-(trifluoromethyl) phenoxy] pyridin-3-yl} amino)-5H-pyrrolo[3,2-d] pyrimidin-5-yl] ethoxy} ethanol) has interacting residues with the protein GLU 770, MET 774, LEU 785, VAL734, LYS 753, THR 862, ALA 751, GLN 799, MET 801, LEU 852, LEU 726, PHE 864, LEU 796, ASP 863, GLY 729, ASN 850.

The crystal structure of the predocked protein is downloaded and docked with the inbound to get affinity energy and coordinates of the inbound ligand, where it is docked on the protein. Further, when other ligands are docked, their binding energy and coordinates are compared to determine the position of the ligand with respect to the inbound. The more the energy and similar the coordinates higher is the affinity (Mukherjee et al., 2010).

The ligand that has the highest binding affinity as far as the study and has a favourable toxicity profile is PubChem Id: 164676, it has binding affinity of -10 (kcal/mol), and has interacting residues of LEU 726, VAL 734, LYS 753, ALA 751, LEU 852, CYS 805, THR 798, ARG 849, GLY 729, ASP 863, THR 862. These residues are active site binding residues forming pi, pi-sigma, alkyl and van der walls interactions. The ligand show a favorable toxicity profile obtained from ProTox analysis, compared to that of control ligand.

5. CONCLUSION

In the early phases of drug discovery, computational identification of anti-cancer natural bioactive compounds is a game-changing method that offers major benefits in terms of speed, cost-effectiveness, and the capacity to screen large chemical libraries. The anti-cancer potential of natural compounds, including databases from traditional medicine, microalgae, and large compound libraries, can be effectively predicted and assessed by researchers using techniques like molecular docking, virtual screening, ADMET analysis, QSAR studies, and molecular dynamics simulations.

This project has shown that in silico methods are very good at identifying promising candidates, offering a time and money-efficient solution that is less harmful to animals than the conventional drug discovery process. The significance of natural products in contemporary oncology drug development is highlighted by the discovery of thousands of compounds in traditional Chinese medicine databases that exhibit a high degree of similarity to approved anti-cancer medications (Dai, SX., Li, WX., Han, FF. et al. 2016). Similar to this, lead compounds from microalgae and other natural sources have been identified through virtual screening and molecular dynamics simulations. These compounds have strong binding affinities and favorable pharmacokinetic profiles, which support their potential as therapeutics for melanoma and other cancers (Prasetiya, Destiarani, et al. 2023).

Biological activity, safety, and efficacy in vitro and in vivo must be confirmed by rigorous experimental validation in addition to computational predictions, despite these advancements (Chunarkar-Patil, Pritee et al., 2024). The accuracy and productivity of anti-cancer compound discovery could be further improved in the future by combining traditional computational methods with artificial intelligence and machine learning protocols.

To sum up, computational identification is a strong and essential method for locating new anti-cancer substances in natural sources. These approaches speed up the process from compound discovery to clinical application by bridging the gap between conventional knowledge and contemporary drug development, ultimately aiding in the creation of safer, more potent cancer treatments.

6. APPENDIX

1. Data Sources and Compound Libraries

ZINC Database: Screened >230 million commercially available compounds for virtual high-throughput screening (vHTS) .

Traditional Medicine Databases: Curated 50,000+ natural compounds from Chinese and Ayurvedic medicinal repositories.

GDB-13: Utilized 970 million synthetically accessible molecules for structure-based drug design .

Microalgae Bioactives: Compiled 2,000+ terpenoids and alkaloids from marine bioprospecting studies.

2. Computational Workflow

Generative AI Models: LSTM Variational Autoencoder (VAE) architecture generated novel compounds with optimized binding properties.

SMILES-encoded sequences trained on FDA-approved anti-cancer drugs.

Structure-Based Approaches: Molecular Docking: Performed with Glide, DOCK, and Surflex for binding affinity prediction

Molecular Dynamics (MD): Simulated ligand-protein interactions for 100 ns to assess stability.

3. Validation Protocols

Binding Affinity Thresholds: Compounds with docking scores ≤ -8.0 kcal/mol prioritized.

ADMET Predictions: Filtered candidates using SwissADME for bioavailability and Lipinski's Rule of Five

7. REFERENCES

Aertgeerts, Kathleen et al. Structural Analysis of the Mechanism of Inhibition and Allosteric Activation of the Kinase Domain of HER2 Protein. Journal of Biological Chemistry, Volume 286, Issue 21, 18756 – 18765

Alsuhaibani, A., et al. (2022). Integrated approach for studying bioactive compounds from *Cladosporium* spp. *Scientific Reports*, 12(1). https://doi.org/10.1038/s41598-022-22038-x

Bakchi, B., Krishna, A. D., Sreecharan, E., Ganesh, V. B. J., Niharika, M., Maharshi, S., ... & Shaik, A. B. (2022). An overview on applications of SwissADME web tool in the design and development of anticancer, antitubercular and antimicrobial agents: a medicinal chemist's perspective. *Journal of Molecular Structure*, *1259*, 132712.

Baroroh, U., Biotek, M., Muscifa, Z. S., Destiarani, W., Rohmatullah, F. G., & Yusuf, M. (2023). Molecular interaction analysis and visualization of protein-ligand docking using Biovia Discovery Studio Visualizer. *Indonesian Journal of Computational Biology (IJCB)*, *2*(1), 22-30.

Biesiada, J., Porollo, A., Velayutham, P., Kouril, M., & Meller, J. (2011). Survey of public domain software for docking simulations and virtual screening. *Human genomics*, *5*, 1-9.

Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* **7**, 42717 (2017). https://doi.org/10.1038/srep42717

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, *7*(1), 42717.

Dar, A. M. and Mir, S. (2017). Molecular docking: approaches, types, applications and basic challenges. Journal of Analytical &Amp; Bioanalytical Techniques, 08(02). https://doi.org/10.4172/2155-9872.1000356

David, B., et al. (2020). Anticancer drug discovery based on natural products. *PMC*. https://doi.org/10.3390/molecules25010156

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*, *40*(1), 82-92.

Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). Autodock vina 1.2.0: new docking methods, expanded force field, and python bindings. Journal of Chemical Information and Modeling, 61(8), 3891-3898. https://doi.org/10.1021/acs.jcim.1c00203

Egan, W. J., Merz, K. M., & Baldwin, J. J. (2000). Prediction of drug absorption using multivariate statistics. *Journal of medicinal chemistry*, *43*(21), 3867-3877.

Elokely, K. M., & Doerksen, R. J. (2013). Docking challenge: protein sampling and molecular docking performance. *Journal of chemical information and modeling*, *53*(8), 1934-1945.

Elyashberg, M., & Argyropoulos, D. (2020). Computer-assisted structural elucidation of natural products. *Frontiers in Pharmacology*. https://doi.org/10.3389/fphar.2021.702611
Harn, N., et al. (2017). NP-Structure predictor for natural product mixtures. *Journal of Natural Products*.

Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of combinatorial chemistry*, *1*(1), 55-68.

Harn, Y., Su, B., Ku, Y., Lin, O. A., Chou, C., & Tseng, Y. J. (2017). Np-structurepredictor: prediction of unknown natural products in plant mixtures. Journal of Chemical Information and Modeling, 57(12), 3138-3148. https://doi.org/10.1021/acs.jcim.7b00565

Huang, M., Lu, JJ. & Ding, J. Natural Products in Cancer Therapy: Past, Present and Future. *Nat. Prod. Bioprospect.* **11**, 5–13 (2021). https://doi.org/10.1007/s13659-020-00293-7

Iqbal, D., et al. (2022). In silico exploration of *Foeniculum vulgare* against breast cancer. *PMC*. https://doi.org/10.3390%2Fijms23137212

Khalifa, S. A. M., et al. (2019). Fragment-based screening of natural anticancer compounds. *Frontiers in Pharmacology*.

Kondapuram, S. K., Sarvagalla, S., & Coumar, M. S. (2021). Docking-based virtual screening using PyRx Tool: autophagy target Vps34 as a case study. In *Molecular docking for computer-aided drug design* (pp. 463-477). Academic Press.

Li, J. (2020). Fragment-based drug discovery. *Journal of Medicinal Chemistry*.
Liu, Y., & Quinn, R. J. (2019). FBDD in natural product discovery. *Current Opinion in Biotechnology*.

Lipinski, C. A. (2004). Lead-and drug-like compounds: the rule-of-five revolution. *Drug discovery today: Technologies*, *1*(4), 337-341.

Muegge, I., Heald, S. L., & Brittelli, D. (2001). Simple selection criteria for drug-like chemical matter. *Journal of medicinal chemistry*, *44*(12), 1841-1846.

Nan, Y., Su, H., Zhou, B., & Liu, S. (2023). The function of natural compounds in important anticancer mechanisms. *Frontiers in oncology*, *12*, 1049888. https://doi.org/10.3389/fonc.2022.1049888

Nugroho, A., & Morita, H. (2019). Structural elucidation challenges in natural products. *Frontiers in Pharmacology*.

O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of cheminformatics*, *3*, 1-14.

Ounthaisong, U., & Tangyuenyongwatana, P. (2017). Cross-docking study of flavonoids against tyrosinase enzymes using PyRx 0.8 virtual screening tool. *TJPS*, *41*(2017).

Pantel, K., Alix-Panabières, C., & Riethdorf, S. (2009). Cancer micrometastases. *Nature reviews Clinical oncology*, *6*(6), 339-351.

Pirhadi, S., Sunseri, J., & Koes, D. R. (2016). Open-source molecular modeling. *Journal of Molecular Graphics and Modelling*, *69*, 127-143.

Priyanka Banerjee, Emanuel Kemmler, Mathias Dunkel, Robert Preissner, ProTox 3.0: a webserver for the prediction of toxicity of chemicals, *Nucleic Acids Research*, Volume 52, Issue W1, 5 July 2024, Pages W513–W520, https://doi.org/10.1093/nar/gkae303

Rinschen, M. M., et al. (2019). Hyphenated techniques in natural product analysis. *Analytical Chemistry*.

Rinschen, M.M., Ivanisevic, J., Giera, M. *et al.* Identification of bioactive metabolites using activity metabolomics. *Nat Rev Mol Cell Biol* **20**, 353–367 (2019). https://doi.org/10.1038/s41580-019-0108-4

Rudmann D. G. (2013). On-target and off-target-based toxicologic effects. *Toxicologic pathology*, *41*(2), 310–314. https://doi.org/10.1177/0192623312464311

Sarkar, A., Concilio, S., Sessa, L., Marrafino, F., & Piotto, S. (2024). Advancements and novel approaches in modified AutoDock Vina algorithms for enhanced molecular docking. *Results in Chemistry*, *7*, 101319.

Sisodiya, D., Pandey, P., & Dashora, K. (2012). Drug designing softwares and their applications in new drug discover. *Journal of Pharmacy Research*, *5*(1), 124-126.

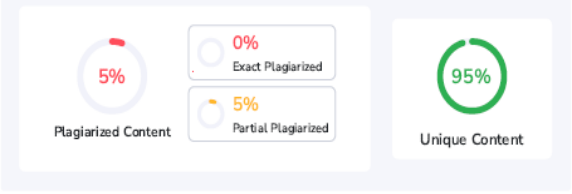
Torre, L. A., Siegel, R. L., Ward, E. M., & Jemal, A. (2016). Global cancer incidence and mortality rates and trends—an update. *Cancer epidemiology, biomarkers & prevention*, *25*(1), 16-27.

Yuan, S., Chan, H. C. S., & Hu, Z. (2017). Using pymol as a platform for computational drug design. WIREs Computational Molecular Science, 7(2). https://doi.org/10.1002/wcms.1298

PLAGARISM REPORT

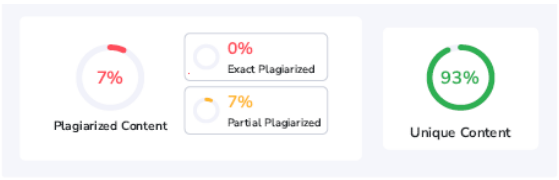
Total Words: 847 Total Characters: 6032 Plagiarized Sentences: 1.17 Unique Sentences: 37.83 (97%)

Total Words: 529 Total Characters: 3604 Plagiarized Sentences: 0 Unique Sentences: 24 (100%)

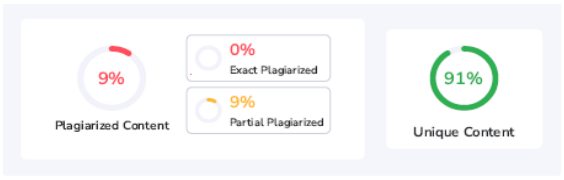


Total Words: 878 Total Characters: 5920 Plagiarized Sentences: 2.05 Unique Sentences: 38.95 (95%)

Total Words: 995 Total Characters: 7040 Plagiarized Sentences: 3.36 Unique Sentences: 38.64 (92%)



Total Words: 674 Total Characters: 4755 Plagiarized Sentences: 2.1 Unique Sentences: 27.9 (93%)



Total Words: 989 Total Characters: 7159 Plagiarized Sentences: 4.05 Unique Sentences: 40.95 (91%)